Final report for the pilot sampling project

Development of a sampling methodology to create an inventory for non-native weeds, within the northern range of Yellowstone National Park.

Principal Investigators:
Lisa Rew, Bruce Maxwell, Richard Aspinall, T. Weaver
and Don Despain

November 2001

Contracted Project Plan

The objective of the six-month pilot project was to determine the most efficient and accurate approach for creating an inventory of the non-native plant species within the northern elk winter range of Yellowstone National Park. The area is too large to look for non-native species over the entirety so we focused on identifying sampling methods that provide the highest probability of locating even the rarest non-native plant species. This objective was achieved through computer simulation and field sampling. Computer simulations of hypothetical plant population distributions were combined with a variety of sampling strategies to evaluate the most efficient sampling method to record species occurring at low frequency within a heterogeneous environment. The chosen sampling method was then used in the field to record the presence of nine selected non-native weeds. Data analysis provided information on the frequency of the non-native weeds within the Park and determined whether their presence was correlated with recorded environmental variables or human activities (i.e. distance from roads and trails). The sampling methodology was subjectively evaluated for practical usability in the field and efficiency of sampling in the Park environment. This work has met the deadline of being collated and presented by December 2001. The information will then be used to develop a sampling strategy for the following three years and submitted as a study plan proposal by January 31st 2002.

Background

The National Park Service is required by law to keep the 34 million ha of parks classified as "natural areas" under their jurisdiction as "unaltered by human activities as possible" (U.S. National Park Service, 1996). The definition of non-native is "any animal or plant species that occurs in a given location as a result of direct, indirect, deliberate or accidental actions by humans" (U.S. National Park Service, 1996). This definition permits the user to recognize and distinguish between changes to animal and plant distributions caused by natural processes and human influences. In reality this statement needs some further clarification. "Human influence" really refers to disturbance by white settlers, more so in the past century and most specifically in the last 50 years.

Many countries have designated specific areas as "wilderness" or "natural ecosystems" and seek to preserve these in their "pristine" state, however pristine is defined. Taking this desire to "protect and retain" such areas, one can argue from the ecological purist point of view, that all non-native species should be removed. However, this is currently impossible from a practical standpoint. In most cases we do not know which non-native species are present within an ecosystem, their frequency or their distribution pattern; how much their distribution is changing and finally what impact they are having on the endemic ecosystem. It is only armed with all of this information that land managers should begin to control or manage non-native species.

The language used to describe the presence and impact of non-native species is often very emotive: "aggressive non-native plants, which spread quickly into natural areas replacing native flora and reducing habitat for native flora and fauna". Often the simple presence of a non-native species is stated as proof enough of present or future environmental damage, particularly if it is a highly competitive species and/or if the increase in the non-native species is associated with the decline of native species. A number of studies have shown that when non-native species are introduced to environments and ecosystems different from those in which they evolved, they may disrupt the ecosystem processes and alter biological diversity (e.g. Braithwaite & Lonsdale, 1989; Hobbs & Mooney, 1991; see Davis *et al.*, 2000 and Mack *et al.*, 2000 for reviews). Invasion by a new species is influenced by three factors:

1. ecosystem properties, which could be related to the level or frequency of disturbance;
2. number of propagules entering a new environment (propagule pressure); and,
3. the properties of the invading species (Lonsdale, 1999).

Davis *et al.* (2000) offer a new theory, that the fluctuation of resource availability is a key factor in controlling invasion. This theory allows for the integration of resource availability with disturbance and fluctuating environmental conditions.

Disturbance is often suggested as a key factor in enhancing the probability of non-native plant establishment in native plant communities. Natural disturbance has a variety of biotic and geomorphic causes including soil disturbance by fauna, weather related events such as mudflows, floods, wind, fire and geological events such as landslides. Fire is sometimes a quasi-human disturbance if management practices suppress, contain or intentionally ignite them, or if fires are ignited accidentally or intentionally by vandals, whichever way, the natural occurrence of fires has usually been altered. Human disturbance includes, construction and use of roads and trails, buildings, utility corridors and campgrounds.

As stated above, the National Park Service has a mandate to preserve the natural systems under their control (National Park Service Organic Act of 1916). There are several phases necessary to achieve this objective:

- Phase 1 creating an inventory (documenting occurrence);
- Phase 2 monitoring (quantifying changes in distribution or abundance); and,
- Phase 3 control or management of non-native species.

To a certain extent these phases can be performed concurrently. The aim of the current project is Phase 1, development of an inventory program.

Pilot sampling study

Yellowstone National Park (YNP) covers an area of 899121 ha. Approximately 1265 plant species have been recorded in YNP of which 187 (15%) are non-native plant species

(Whipple, 2001). The pilot sampling study concentrated on the area within the northern elk winter range of the Park (152785 ha).

The relative proportional importance of the different forms of disturbance on non-native species establishment and survival has not been quantified. The general perception from the Yellowstone National Park staff involved with weed surveys and members of this research group was that most of the infestations occur close to roads, trails and human habitation. From the data previously collected by park staff in 1998, it was calculated that 278 of 422 (66%) weed occurrences were less than 100 m from roads or trails, and all observations were made less than 500 m from roads or trails. These data were not collected using a formal sampling strategy and the sites searched were biased by their proximity to roads and trails. Therefore, this information was treated as anecdotal and although considered, the data were not used for any subsequent analysis.

Simulation of sampling methodologies

To ensure the best use of the limited funds and time available in the field, a desktop study was conducted to develop the most effective sampling regime. This was performed in ESRI ArcView© GIS using a routine developed by R. Aspinall. This implemented several different sampling strategies including simple random sample, random walk, random transects, transects normal to specified linear features, stratified random sampling, and regular (grid) sampling. Additionally, different sampling intensities were evaluated for different infestation levels (frequencies) of non-native plants.

Using equation 0.1 it is possible to use a random sample to determine the percentage of the study area occupied by non-native plants.

$$p = \left(\frac{100n}{N} \right) \tag{0.1}$$

Where:
p = the percentage of the study area occupied by non-native plants,
n = the number of sample points at which the non-native plant is found, and
N = the total number of sample points searched.

The same relationship applies to samples along lines (transects) in which case n is the length of transects occupied by the target non-native plant, and N is the total transect length searched.

Working through an example of this, within ArcView we simulated 100 non-native plant patches, each 3 m long. 100 transects, each 2000 m in length, are randomly generated.

ArcView is then used to identify the length of the transects that are occupied by non-native plants (by intersecting the 3m non-native plant patches with the 100 transects). The length of transect occupied by non-natives is 300m. Inserting this in equation 0.1 with $n = 300$, $N = 200000$ gives $p = 300/200000$ or 0.15%. In general, with any random sampling scheme the target species is sampled at the frequency with which it occurs in the landscape; this applies to any distribution pattern. Note that, for sampling, frequency is measured as the **area** occupied by the species of interest and not as the level of abundance of the species.

It is assumed that most of the species we are targeting are at a low frequency within the landscape and therefore collecting large numbers of observations is important to provide a reliable estimate of the species occurrence. A large sample combined with an appropriate strategy for estimating geographic distribution is also necessary if the goal is to estimate the distribution of the non-native plant in the landscape. Survey design is, therefore, a tradeoff between collecting a sufficiently large sample to provide reliable estimates of occurrence, and using a sampling strategy that is efficient for both a) field work and b) estimating the geographic distribution of the species.

The simulations and sampling strategies implemented within the GIS allows us to evaluate which sampling strategy provides us with the highest number of sample points for the shortest time in the field and that also provides geographic coverage necessary for estimating distribution of the non-native species. Random points, for example, are not as efficient for collecting data as random walks or transects since time used moving from one survey location to another location is not used for data collection. Surveying along transects allows data to be collected continuously and a large sample size be generated. Additionally, surveying along transects allows changes in underlying environmental variables to be recorded. This is important for estimating the geographic distribution of the species from the sample data.

If the occurrence of a target species is known to be correlated with an environmental variable, we could stratify the sampling scheme on that variable and improve our probabilities of finding the target (Hirzel and Guisan, in press). We accepted the assumption that human disturbance in the form of roads and trails increases the chance of finding non-native weeds, and stratified our sampling using this variable. However, to test this hypothesis we also need to sample away from roads and trails. Therefore, transects established perpendicular to roads and trails were accepted as the most effective sampling methodology. The use of 2000 m transects allows the importance of other factors to be evaluated, since each transect is sufficiently long to cross a number of habitat types and other environmental transitions.

Collection of field data

Four sub-areas were selected from the northern range for the field pilot study. These are an intensively disturbed area (near Mammoth) and three less disturbed areas around Blacktail Creek, Tower and Lamar Valley (Fig. 1). Within each of these areas the starting position of each transect was randomly selected along a road or trail and constructed perpendicular to the road or trail. This was performed using ArcView and prior to arriving at the site.

In the field, transects were walked by two person teams. Transects were 10 m wide. Locations were recorded with a GPS (Global Positioning System). Trimble Pro XRS receivers and occasionally a GeoExplorer® 3 were used and the data post-processed using differential correction to improve accuracy. Manufacturers quotation for post-processing differential correction horizontal accuracy is sub-meter for the XRS receiver and 1 – 5 meter for the GeoExplorer® 3. We recorded average and median horizontal precision values of 0.722 m and 0.629 m respectively for the XRS receiver and 0.820 m and 0.652 m respectively for the GeoExplorer® 3. A point worth noting was that the GeoExplorer® 3 recorded "no service" more frequently than the XRS receivers and therefore caused a loss of data collection. The positioning of the GeoExplorer® 3 antenna was also more sensitive than the XRS receiver antenna. The coordinate system and projection used for the GPS was Universal Transverse Mercator (UTM) Zone 12N, WGS 1984 Datum. This projection and datum is the same as used for GIS data maintained by YNP Center for Resources, and the Greater Yellowstone Area Spatial Data Clearinghouse managed and maintained by the Geographic Information and Analysis Center (GIAC) at Montana State University.

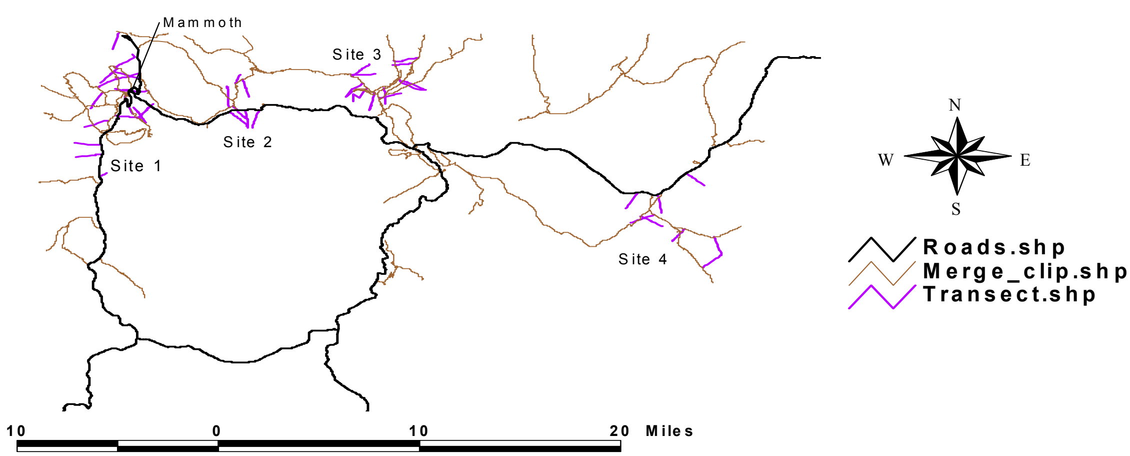
The non-native species targeted for the preliminary study are listed below (Table 1). These were chosen as they were believed to represent the range of non-native plant frequencies and detectabilities found in the Park. The location of any of the target species was recorded concurrently with key environmental variables including habitat type, aspect, topography and disturbance. In addition, whenever the habitat changed or a human disturbance was reached, but no target species were present, the location was again recorded along with the other environmental variables. This provides data to measure the length of transects occupied by non-native species well as the distribution of habitat types, aspect, topography and disturbance along the transects.

Table 1. List of the non-native species used as targets in the field-sampling program.

Distribution: Ease of observation	**Widespread**	**Scattered**	**Local**

Easy	Dalmatian toadflax (*Linaria dalmatica*)	Yellow sweetclover (*Melilotus officinalis)*	Ox-eye daisy (*Chrysanthemum leucanthemum*)
Moderate	Timothy (*Phleum pratense*)	Cheatgrass (*Bromus tectorum*)	Spotted knapweed (*Centaurea maculosa*)
Hard	Canada thistle (*Cirsium arvense*)	Smooth brome (*Bromus inermus*)	Houndstongue (*Cynoglossum officinale*)

Fig.1 Location of the four sites selected for sampling the occurrence of nine non-native plant species and position of the transects walked.



Analysis of northern range field data

Over the field season 42 transects were walked, 16 near the Mammoth area, and 7, 11 and 8 at the Blacktail Creek, Tower and Lamar Valley sites respectively. The average time taken to walk and record a transect was 2 hours, but this did not include walking to or from the start/finish point back to the parking lot etc. Not all transects were exactly 2000 m long due to difficulties with the terrain and inaccuracies with non-differential GPS. The overall length sampled was 86 053 m (x 10 m wide). The total number of target non-native plant infestations observed was 1055. Using equation 0.1 above and assuming that the average length of patches was 3 m, we calculated the proportion of the of the study area infested with each of the target species.

Table 2. Total number of observations and the percentage of the study area infested by target species, as calculated using a total transect distance of 86053 m and assuming a patch size of 3 m in length.

Species: % of area infested	Timothy	Canada thistle	Cheat grass	Smooth brome	Dalmatian toadflax	Yellow sweet-clover	Ox-eye daisy	Spotted knapweed	Hounds-tongue
All sites	1.618	0.377	0.262	0.272	0.561	0.314	0.028	0.038	0.209
Mammoth only	1.868	0.596	0.434	0.515	1.424	0.879	0.061	0.111	0.606
Other sites	1.486	0.261	0.170	0.144	0.107	0.016	0.011	0.000	0.000
Total observations (n)	464	108	75	78	161	90	8	11	60

When all sites were combined the percentage of the study area infested was above 0.2% for all species except ox-eye daisy (*Chrysanthemum leucanthemum*) and spotted knapweed (*Centaurea maculosa)*. However, infestations were perceived to be higher around Mammoth which could skew the results. Percentage infestations were calculated separately for Mammoth and the three remaining sites (Table 2). The percentage infestation for the 16 transects around Mammoth were relatively high; Dalmatian toadflax (*Linaria dalmatica*) and timothy (*Phleum pratense*) were present at levels of 1.4 and 1.9 % respectively, the remaining species were present at densities above 0.4%, with the exception of ox-eye daisy and spotted knapweed. When considering only the Blacktail, Tower and Lamar sites the percentage infestation of timothy was still relatively high (1.5%) but the values for all the other species were less than half those calculated for Mammoth.

Graphical display and analysis of the target species data against environmental data was performed using the data we collected in the field, rather than environmental data previously developed in GIS. The resolution of the GIS data is 30 m for the DEM and 50 m for the habitat and geology data; these resolutions are considered too coarse to analyze environmental associations of the non-native plants for the pilot study. Dalmatian toadflax, houndstongue (*Cynoglossum officinale*) and yellow sweetclover (*Melilotus officinalis)* were more common on an easterly aspect (67.6 – 112.5°). Cheatgrass (*Bromus tectorum*) was more common on the easterly and southerly aspects (67.6 – 112.5° and 157.6 – 202.5° respectively) (Fig 2). Timothy, Canada thistle (*Cirsium arvense*) and smooth brome (*Bromus inermus*) were equally prevalent on all aspects (Fig. 2). Too few data points were collected for ox-eye daisy and spotted knapweed to determine a pattern. All the species, except timothy, occurred more frequently on sites with no aspect, i.e. level areas, but it should be considered that this information is confounded with moisture and angle of slope factors, as level areas generally also indicated increased moisture availability. Data were weighted to allow differences to be detected when there were different numbers of observations recorded for each aspect.

The equation used to weight the data is provided below:

$$P = \left(\frac{n_x}{\sum n}\right) \cdot \left(\frac{n_{ix}}{\sum i}\right) \qquad (0.2)$$

Where:
P = proportion,
n = number of observations,
x = variable of interest (habitat, distance, etc.), and
i = target species of interest.

Graphical display of the habitat and target species data suggests that the big sage /bluebunch wheat grass (*Artemisia tridentata/Agropyron spicatum*), big sage/Idaho fescue (*Artemisia tridentata/Festuca idahoensis*) and to a lesser extent douglas fir/pinegrass *(Pseudotsuga menziesii/Calamagrostis rubescens*) habitats had a higher proportion of the target species (Fig. 3), even when the data were weighted to allow for the higher numbers of observations taken in these habitats.

The distance from road/trail data for the transects was re-calculated to allow for the fact that while transects initially were perpendicular to the road or trail at the point at which they commenced, this does not mean that transects monotonically increased in distance from any road or trail. The distance to trails and roads was re-calculated for the transects and the distance partitioned into 100 m intervals; these data are shown in Fig. 4. The proportion of observations made at 100 m intervals was used to weight the proportion of each target species observed within those intervals. For all species there is a very marked decline after 100 m from road and trails (Fig. 4).

Fig 3. Proportion of different aspects sampled and the weighted proportion of each target species at the different aspects.

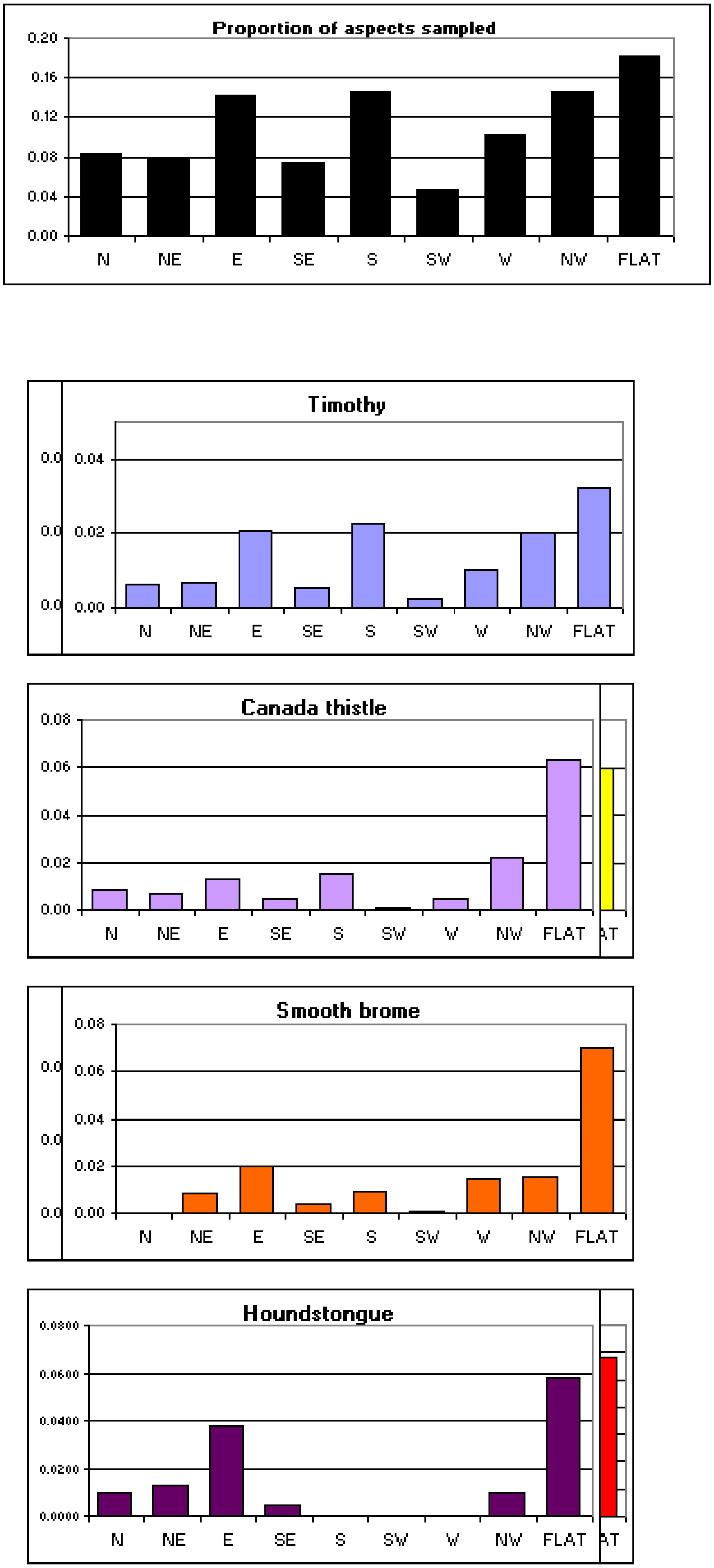


Table 3. List of common, scientific and abbreviated code names for the habitats recorded in the study area.

Common names	***Scientific names***	Abbreviated name
Big Sagebrush/Bluebunch Wheatgrass	*Artemisia tridentata/Agropyron spicatum*	artr/agsp
Big Sagebrush/Idaho Fescue	*Artemisia tridentata/Festuca idahoensis*	artr/feid
Big Sagebrush/Idaho Fescue-Sticky Geranium Phase	*Artemisia tridentata/Festuca idahoensis-Geranium viscosissimum*	artr/feid/gevi
Douglas Fir/Bluebunch Wheatgrass	*Pseudotsuga menziesii/Agropyron spicatum*	psme/agsp
Douglas Fir/Common Juniper	*Pseudotsuga menziesii/Juniperus communis*	psme/juco
Douglas Fir/Heartleaf Arnica	*Pseudotsuga menziesii/Arnica cordifolis*	psme/arco
Douglas Fir/Idaho Fescue	*Pseudotsuga menziesii/Festuca idahoensis*	psme/feid
Douglas Fir/Pine Grass	*Pseudotsuga menziesii/Calamagrostis rubescens*	psme/caru
Douglas Fir/Snowberry	*Pseudotsuga menziesii/Symphoricarpos albus*	psme/syal
Douglas Fir/Snowberry, Big Sagebrush/Idaho Fescue, Sticky Geranium Phase	*Pseudotsuga menziesii/Symphoricarpos albus, Artemisia tridentata/Festuca idahoensis-G. viscosissimum*	psme/syal/artr/ feid/gevi
Idaho Fescue/Agropyron sp.	*Festuca idahoensis/Agropyron sp.*	feid/agsp
Idaho Fescue/Agropyron sp.-Sticky Geranium Phase	*Festuca idahoensis/Agropyron sp.-Geranium viscosissimum*	feid/agsp/gevi
Idaho Fescue/Richardson's Needlegrass	*Festuca idahoensis/Stipa richardsonii*	feid/stri
Lodgepole Pine	*Pinus contorta*	pico
Lodgepole Pine/Elk sedge	*Pinus contorta/Carex rossii*	pico/caro
Sedge bogs	*Carex sp. bogs*	Carex sp.
Silver Sage/Idaho Fescue	*Artemisia cana/Festuca idahoensis*	arca/feid
Spruce/Common Horsetail	*Picea sp./Equisetum arvense*	Picea sp./eqar
Willow/Sedge	*Salix sp./Carex sp.*	Salix sp./carex

Fig 3. Proportion of different habitat types sampled and the weighted proportion of each target species in different habitats. See Table 3 for habitat definitions.

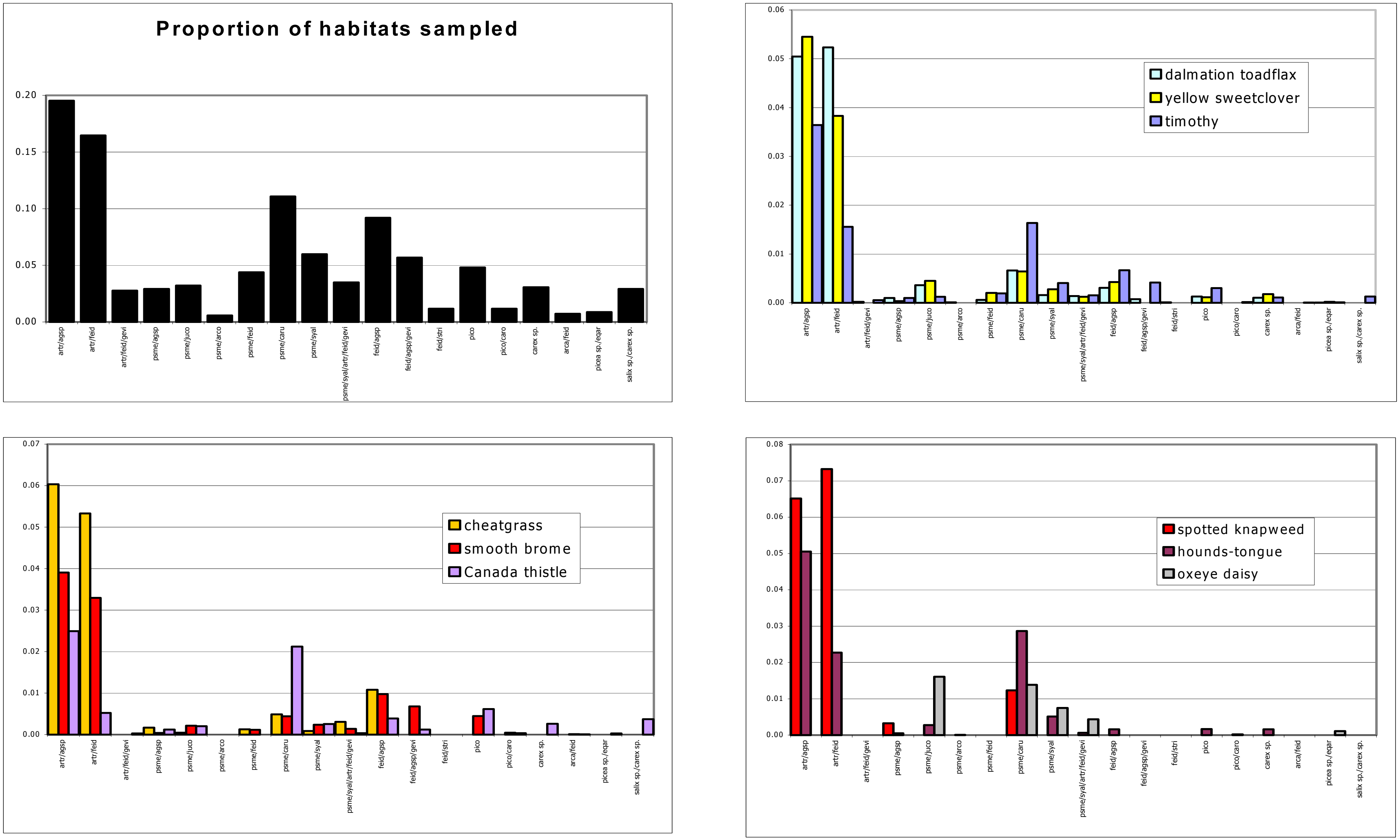
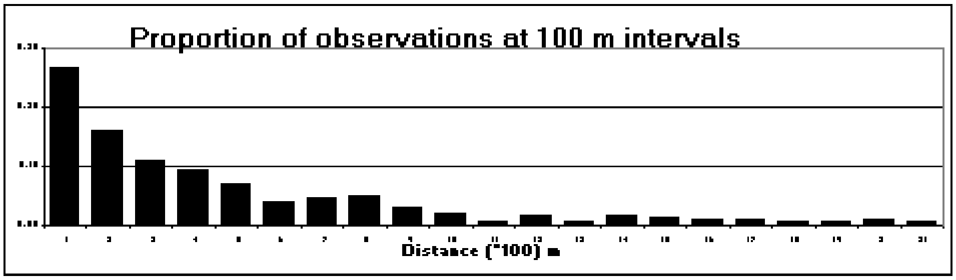
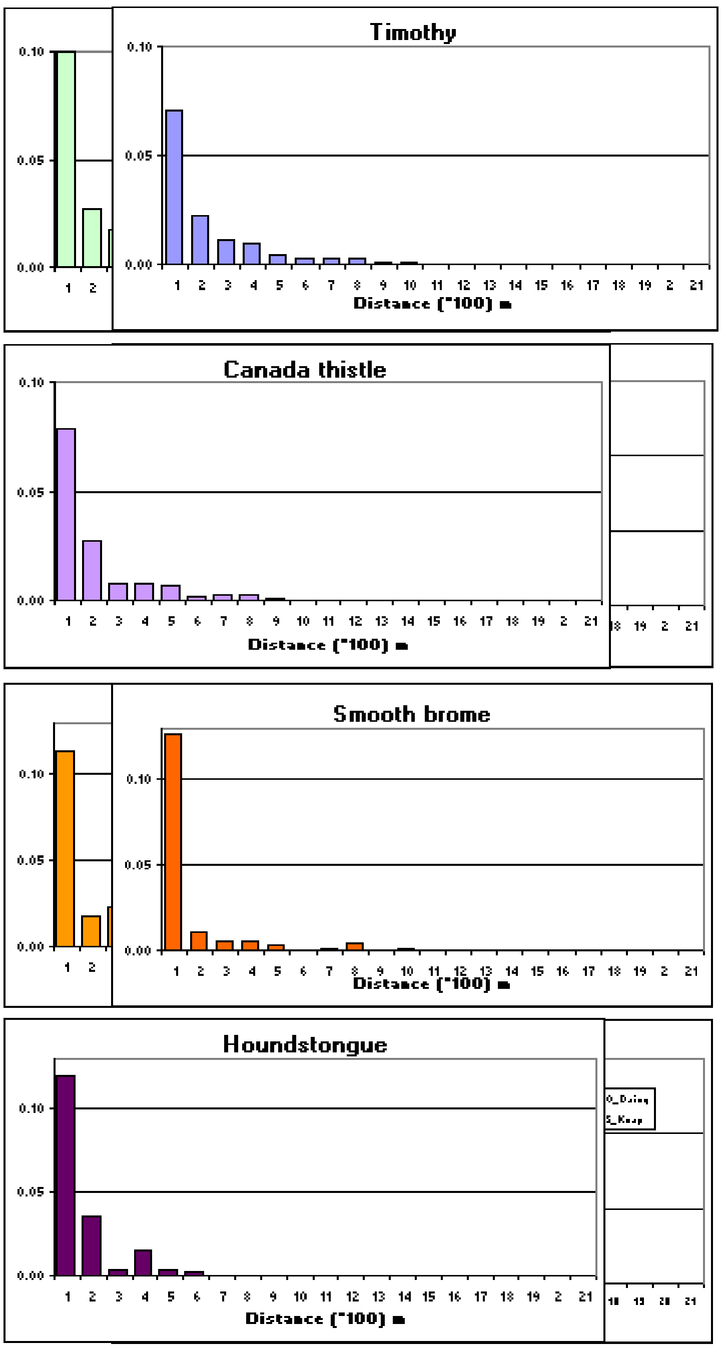


Fig 4. Proportion of observations within 100m intervals of roads/trails and the weighted proportion each target species at 100m intervals from trail/roads.





The data suggest that distance from road/trail, habitat and aspect may all be affecting the occurrence of the nine target species. To determine whether, any of the environmental variables were accounting for more of the variance than any other, that is having more of an "influence" on the target species occurrence data, we used principle component analysis (PCA). This is a particularly useful technique in cases such as this, where there are a large number of observed variables. As the method is not scale invariant and the values for the different variables were on a different scale, correlation matrices were used. Variables used included, aspect, habitat type, distance from road/trail and species occurrence. PCA essentially calculates the proportion of variance that is accounted for by each of the variables. The analysis showed the proportion of variance accounted for by the different variables was similar, suggesting that none of the variables we measured were having a greater impact on the data than any other.

The environmental data displayed above were collected in the field. If the intention is to stratify the sampling by environmental variables it needs to be using data that has already been collected or collated, preferably in GIS format. We therefore used GIS data layers available from the Park. The available habitat and geology data are at a 50 m resolution (though the smallest sampling size was 4 ha) and the DEM at a 30 m resolution; aspect is derived from the 30m DEM. These data were used with the target species data and analyzed with an inductive modeling procedure based on Bayes' theorem, within ArcView (Aspinall, 1992). The aim of this modeling procedure is to predict the distribution of a dataset containing presence/absence data by combining a number of other data sets. Data set combination is carried out using Bayes' theorem. Inputs to the theorem, in the form of conditional probabilities, are derived from an inductive learning process in which attributes of the data set to be modeled (presence/absence) are compared with attributes of a variety of predictor data sets (see Aspinall, 1992, for more detail).

The associations between the non-native species and field measured environmental data (above) were also apparent in the coarser resolution GIS data. As part of the analysis, the Bayesian modeling method uses the Chi-square test to identify associations between the presence/absence data and an environmental data set that is a potential predictor. Unfortunately, the table of association between the target species and environmental data contained many zeroes (absences) and even where target species were present the numbers were often too low to provide a valid Chi-square test. Larger sample sizes will address this issue.

However, another reason for the lack of significant associations could be related to the scale of pattern of non-native weeds in the landscape. It may be that the scale of pattern displayed for the non-native weeds is less than the resolution of the environmental variables. We computed the three-term local quadrat variance (Hill, 1973) to identify scale of pattern along a transect. Analysis was performed on two transects containing cheatgrass assuming a static patch size of 3 m. The scale was calculated as less than 25 m. More analysis needs to be performed on other species and to allow for the different patch sizes recorded in the field. This analysis needs further development and cannot be included in this report but it will be performed at a later time. The preliminary results do suggest that the scale of spatial pattern in the non-native species data could be much finer than the minimum spatial resolution of the (relatively coarse) GIS data available. If this is the case, then existing GIS data will only be able to provide very general models of non-native species distribution.

Conclusions

The simulation study was a time efficient way to evaluate different sampling methods. Field data collection and analyses have provided us with a better quantitative understanding of the frequency of non-native weeds and shown trends between non-native weed presence and human activity, in the form of roads/trails and environmental variables. Time spent in the field has provided us with better practical and conceptual understanding of the issues involved with sampling within large heterogeneous environments. Stohlgren *et al*. (1995) stated that there are no "off-the-shelf sampling designs and techniques…to create biotic inventories", but using the results we have obtained in this pilot study we will be able to refine and define a scientifically rigorous and valid sampling strategy for the northern elk winter range of Yellowstone National Park, which will be applicable to other natural areas.

Acknowledgements

Software to reformat the field target weed data, and implement different sampling strategies in ArcView, were written by Richard Aspinall.
Thanks to all those who helped with the field sampling, Matt Hulbert, Jeremy Gay, Judit Barosso, Jerad Corbin, Kit Sawyer and Nicole Wagner.

References

Aspinall R (1992) An inductive modeling procedure based on Bayes' theorem for analysis of pattern in spatial data. *Int. J. Geog. Systems*, 6, 105-121.

Braithwaite RW, Lonsdale WM & Estbergs JA (1989) Alien vegetation and native biota in tropical Australia: the spread and impact of *Mimosa pigra*. *Biol Cons*., 48, 189-210.

Davis MA, Grime JP & Thompson K (2000) Fluctuating resources in plant communities: a general theory of invisibility. *J. of Ecology*, 88, 528-534.

Hill MO (1973) The intensity of spatial pattern in plant communities. *J. of Ecology*, 61, 225-235.

Hirzel and Guisan (in press). Which is the optimal sampling strategy for habitat suitability modeling. Ecological Modeling.

Hobbs RJ & Mooney HA (1991) Effects of rainfall variability and gopher disturbance on serpentine annual grassland dynamics. *Ecology*, 72 478-491.

Lonsdale WM (1999) Global patterns of plant invasions and the concept of invisibility. *Ecology*, 80, 1522-1536.

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M & Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.*, 10, 689-710.

U.S. National Park Service 1996 (Online) Preserving our natural heritage – a strategic plan for managing invasive non-native plants on national park system lands. http://www.nature.bps.gov/wv/strat_pl.htm

U.S. National Park Service Organic Act of 1916 (Online) (Exerts from) http://www.nps.gov/legacy/nomenclature.html

Whipple JJ (2001) Annotated checklist of exotic vascular plants in Yellowstone National Park. *West. North Am. Nat.*, 61, 336-346.